Insect-inspired visual navigation for flying robots

Andrew Philippides[1,*], Nathan Steadman[1], Alex Dewar[2], Christopher Walker[1] and Paul Graham[2]

[1] Centre for Computational Neuroscience and Robotics, Department of Informatics, University of Sussex, Brighton, UK
{andrewop,N.Steadman,Chris.Walker}@sussex.ac.uk
[2] Centre for Computational Neuroscience and Robotics, School of Life Sciences, University of Sussex, Brighton, UK
{A.Dewar,paulgr}@sussex.ac.uk

Abstract. This paper discusses the implementation of insect-inspired visual navigation strategies in flying robots, in particular focusing on the impact of changing height. We start by assessing the information available at different heights for visual homing in natural environments, comparing results from an open environment against one where trees and bushes are closer to the camera. We then test a route following algorithm using a gantry robot and show that a robot would be able to successfully navigate a route at a variety of heights using images saved at a different height.

Keywords: Visual navigation, insect-inspired robotics, visual homing, UAVs, flying robots

1 Introduction

Navigation is a vital ability for animals and robots. In the latter, where GPS is unavailable or unreliable, visual homing methods can be used. For flying robots, the limitations on payload mean that it is important that any algorithms are efficient in terms of computation as this reduces power consumption. It is therefore natural that engineers turn for inspiration to insects such as ants and bees [1], who use vision to navigate long distances through complex natural habitats despite limited neural and sensory resources [2-5]. In this spirit, we have previously developed a view-based homing algorithm based on the behavior of desert ants. While this algorithm has been tested for ground-based robots in simulation, here we test the ability of this algorithm to generalize to a flying robot.

The first generation of biomimetic algorithms for insect-like navigation were inspired by the fact that an insect's use of vision for navigation is often a retinotopic matching process (Ants: [2,4,5]; Bees: [3]; Hoverflies: [6]; Waterstriders: [7]; Review: [8]) where remembered views are compared with the currently experienced visual scene in order to set a direction or drive the search for a goal. Insect-inspired robotic models of visual navigation have thus been dominated by snapshot-type mod-

els where a single view of the world, as memorized from the goal location, is compared to the current view in order to drive a search for the goal ([3]; for review see [9]). Importantly, Zeil and colleagues [10,11] used simple metrics based on the sum-square differences in pixel values between images, and thus agnostic of the details of the model used, to analyse the range over which a single image can be used for visual homing.

This work and others showed that, while snapshot models work in a variety of environments, these approaches are limited in that they generally allow for navigation only in the immediate vicinity of the goal [10-13]. We therefore developed a familiarity-based model of route navigation in which individual views are used as a visual compass to recall the direction the agent was facing at that point, rather than the direction to the location as in most snapshot models. This allowed us to develop a holistic route memory which allowed a ground-based agent to navigate through simulated natural environments with route showing many characteristics of ant routes [14]. Following on from this work, similar models have been shown to work with a biologically plausible neural network in simulation [15] and on a robot in a natural environment [16].

In all these previous works, however, all images were taken from the ground level. To assess the whether this model can also be used for a flying robot, we thus need to test the model with images gathered from different heights. To do this we follow the methods of [10,11,17] to first assess the extent over which single images gathered from different heights through two natural environments can provide information for visual homing. We show that in line with Zeil [10], a snapshot stored at one height can successfully be used as either an attractor snapshot or visual compass for an agent travelling at a different height. We then test our route navigation algorithm with similar data gathered using a high precision gantry robot in an indoor environment. The success of the algorithm provides proof of principle that an aerial robot could use our route navigation algorithm despite being at a different height to that at which the original route was travelled.

2 Measuring the informational content of natural images from different heights

To get an understanding of the information that is present in images for homing, we follow the procedures of [10,11,17] and use simple metrics to estimate the range over which a single image can be used either as an attractor type snapshot, to recover a direction towards a goal, or as a visual compass, to recall the heading at which the agent was facing when the goal image was stored.

2.1 Data collection and image processing

The process starts by capturing sets of images using a panoramic imaging device within two natural environments at different heights (Fig. 1). Data was collected using a Kodak Pixpro SP360 camera to take panoramic images at regular intervals at a

range of heights from straight-line routes in two locations. The heights investigated were 0cm, 40cm, 70cm, 100cm, 150cm and 200cm. With the exception of 0cm, when the camera was placed directly upon the ground, images were taken from a tripod with an in-built spirit gauge, allowing each image to be taken at a level setting. In addition, it allowed us to examine the impact of pitch by including an image taken with a 45° downward tilt at each of the heights and locations investigated. The tripod was aligned to a common heading to ensure all images from each location were taken facing in the same direction.

Fig. 1. Sample images from two locations. A: Location A – open transect through Queen's Park, Brighton. B: Location B – transect through wooded copse in East Brighton Park. Images unwrapped from fish-eye lens via PixPro software.

Location A: The initial dataset was collected from a field in Queen's Park, Brighton. These transects covered a distance of 36m taken in 2m increments through an open area. This location was selected as it is an open environment without any nearby trees or foliage and in which the trees which constitute the skyline were at least 50m from the camera location with the majority being several hundred metres away (Fig. 1A). This means the visual data changed slowly relative to movement along the route.

Location B: The second dataset was collected from a small copse in East Brighton Park, Brighton. Transects of 12m were covered in 1m intervals. The treeline was at no point further than 5m from the camera in any of the images taken. The decreased interval between images was used to compensate for the small distance of the route. Due to the forested nature of this site (Fig. 1B), the route passed noticeably between shade and light with distinct lens flare on certain images taken. The images were not manipulated to compensate for this.

The images were processed in three ways: they were first unwrapped from a fish-eye image to a panoramic image using the software provided with the PixPro camera. Images were then scaled-down to 408 x 1632 pixels, converted to grey-scale and the sky homogenized to a uniform value of 255 (white). The first two stages were achieved with Matlab functions imresize and rgb2gray, respectively. Sky homogenization was also performed in Matlab automatically and quite roughly by thresholding the blue-channel at a value of 170 (determined by trial and error) and setting above 170 values to 255. Occasionally, parts of clouds were below the 170 threshold, and so

we also converting any isolated ‘not-sky’ pixels which were not connected to the ground to 255. Sky homogenization is necessary as the light gradient can provide a very strong, but spurious cue, by which homing algorithms can gain information.

2.2 Assessing the region in which an image can be used for visual homing

By measuring the image difference between a reference image and images from surrounding points, we can build an image difference function (IDF) that shows how images change with distance from a goal view [10]. The image difference between two images X and Y is defined as:

$$IDF(X,Y)=\frac{1}{P}\sum_{i}\sum_{j}\left|X(i,j)-Y(i,j)\right| \tag{1}$$

where X(i,j) is the pixel in the i'th row and j’th column of image X and P is the number of pixels. Notice that this value is dependent on the alignment of image X and Y to a common heading. The IDF for an image from Location A is shown in fig. 2.

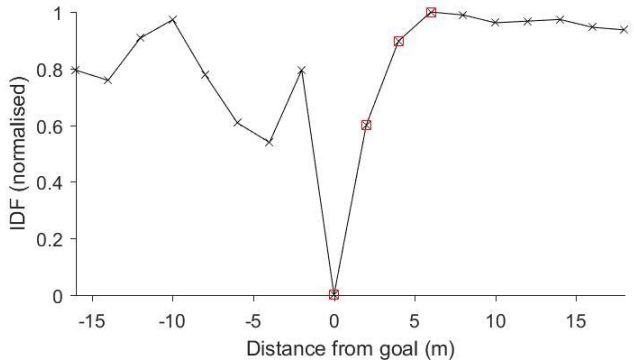

Fig. 2. An example IDF from Location A at a height of 100cm. IDF values are shown as a proportion of the maximum and plotted against distance from the goal. x's mark points where images were taken from. Points within the catchment area are marked with red squares. Noise in the IDF limits the catchment area to the left of the goal while to the right it extends 6 m.

To assess the range over which a single image can be used for navigation, we quantify the region around a goal image in which the difference between images increases with increasing distance from the reference image. The presence of an increasing IDF is significant as it shows that the information needed for view-based homing is available [10,17]. We therefore define the catchment area (CA) of the IDF as the region where the IDF is generally increasing. Note that as we are applying this to a straight line transect this is not strictly an ‘area’ but an indication of the distance over which a single image can be used. Nevertheless we use the term ‘catchment area’ to be consistent with the generally used terminology. To estimate this region, we take

the number of consecutive locations spreading out from the goal position where the IDF is strictly increasing relative to the direction of movement from the goal (Fig. 2). This is a lower bound on the region within which an image could be used for homing as a single a single uneven/bad image can introduce a spurious minimum in this discrete data (e.g. point at distance -2 in Fig 2). While this could be overcome by taking more images and smoothing the data in some way, as we are using the same metric to compare two data sets, we use the raw data as a lower bound. Likewise, the fact that the distribution of catchment areas is asymmetric (goal positions at the centre have CAs that can extend in both directions unlike those at the edge) is not so important as we are comparing the same positions for different heights.

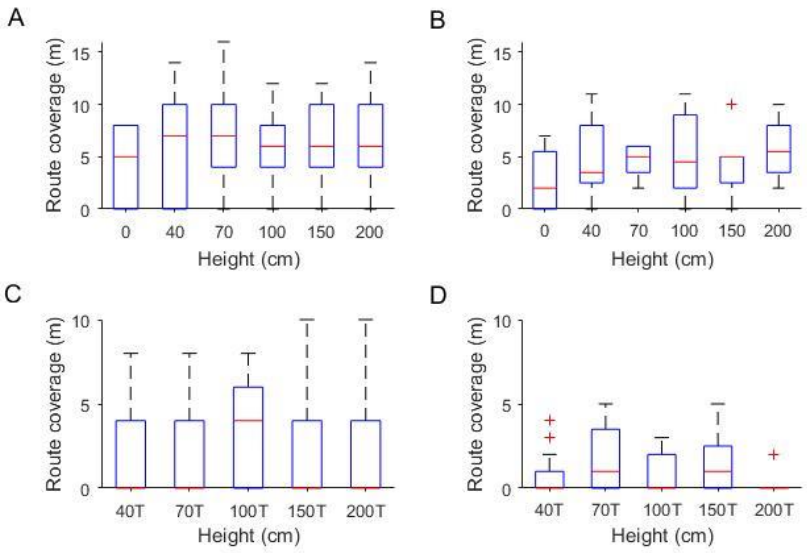

Fig. 3. Catchment areas in locations A (A and C) and B (B and D). X-labels identify the height of the route. In the top row (A and B), data is shown for goal images taken from the route tested. In contrast, height labels followed by a 'T' (bottom row, C-D) indicate that the goal images only were captured at a 45° tilt. Red lines indicate the median catchment area, whole boxes cover the 25th and 75th percentile and whiskers extend to 1.5 times the inter-quartile range.

Fig. 3 A-B shows that images gathered at different heights do carry sufficient information for visual homing. While there is quite a spread of results at each height from different goal positions, a good proportion of the route can be traversed using a single image. As expected, a greater distance could be traversed with a single image in Location A than Location B. Interestingly, in the more cluttered Location B, the performance increased with increasing height. The cause of this can be seen in Fig.4. When there are nearby objects, a limited field of view in the elevation access means that close objects dominate the skyline more for lower heights, leading to inaccuracy in the matching process. We also tested routes taken with a level camera against goal images where the camera was angled at 45º to the ground plane (Fig. 3 C-D). The

idea of this was to see whether the effect of pitch changed with changing height. For this data set, the impact of pitch was detrimental to the route coverage as has been seen in [18] and did not vary with height, but at least in Location A, there remains some information for homing.

Fig. 4. The impact of increased height at Location B. **A** - Location B taken at a height of 40cm. **B** - Location B at a height of 150cm. The increased elevation in B results in a more distinct outline of the treeline to the left of the image. This in turn leads to better IDF comparisons as there is now more visual information.

We next tested whether images that were taken at a particular goal height, could be successfully used to home for a robot travelling at a different height (Fig. 5). Results were mixed but showed that: firstly, in principle this is possible; secondly, and unsurprisingly, that the closer in height the goal and route, the better the match; And finally, and more interestingly, that there was a slight trend of better performance for goals and routes at higher heights. The reasons for the benefits of height are likely to do with the lessening amount of the ground plane in the image as well as a cleaner skyline, though this needs further investigation.

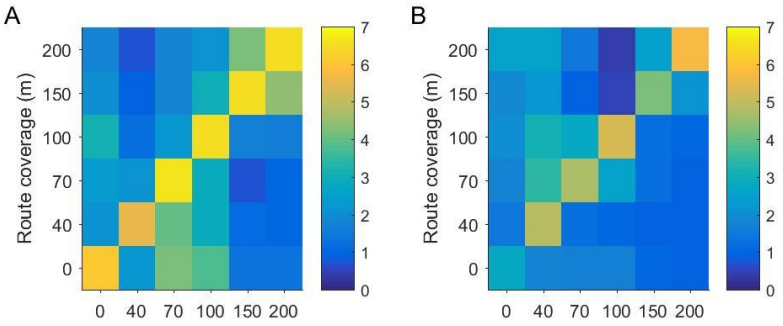

Fig. 5. Catchment areas where the goal (on the x-axis) is at a different height from the route (y-axis). Colours show the mean catchment area across all goal positions in Location A (A) and B (B).

2.3 Assessing the region over which views can be used to recall a heading

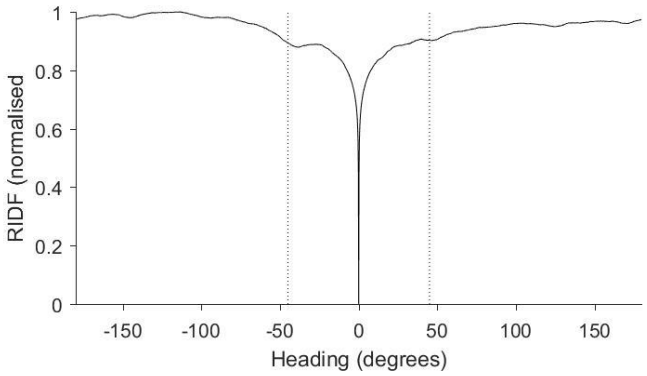

Figure 6: Sample RIDF using an image from Location A, 100cm height, as both goal and current image. IDF is normalised to the maximum value and plotted against the degree of azimuthal rotation. Dashed lines mark the 45° threshold region. In this case the best match is at 0° meaning the agent is facing in the same direction as when the image was stored.

While the analysis of the catchment area of natural images demonstrates that the information required for visual navigation is present, to more directly link it to our familiarity based algorithm we need to repeat the analysis using views as a visual compass. We therefore next determine what we termed rotational catchment areas (RCAs) [17], which indicate the region in which an agent would be able to use a goal image to recall the heading it was facing when the goal images was stored.

To determine the RCA, one first calculates the rotational IDF (RIDF) by evaluating the IDF between a reference image and the current image rotated (in silica) in steps of 1° of azimuth resulting in a 1×360 RIDF (Fig. 6). The minimum value in the RIDF defines an orientation of the current image which gives the closest match with the reference image. In the vicinity of the reference image these orientations will be similar to the reference image orientation [10] and so can be used to recall a heading or a movement direction for a forward facing agent [17]. We define the rotational catchment area as the region spreading out from the location of the reference image where the minimum in the RIDF is less than 45° from the true orientation of the reference image (Fig. 6).

Analysis of RCAs for routes and goals at the same height shows that there are strong visual compass cues across the routes (Fig. 7 A-B) with much increased catchment areas compared to the IDF analysis. In Location A, performance increases with height of the route. In Location B the highest routes seem to slightly underperform but this is likely a ceiling effect (the maximum possible RCS is 11 m for Location B. Comparing goals at different heights to routes again there are large catchment areas for the visual compass with much more matching across different heights than was evident in the IDF data. In Location A, there seem to be two broad blocks of matching with all routes from 100 cm down broadly matching while the two highest

routes match each other. The picture for Location B is similar but with a more broad matching across heights. While the rather crude measure of mean RCA and the ceiling effect for Location B means more investigation is needed, these results bode well for our route following algorithm to function at different heights as it is based on the visual compass method.

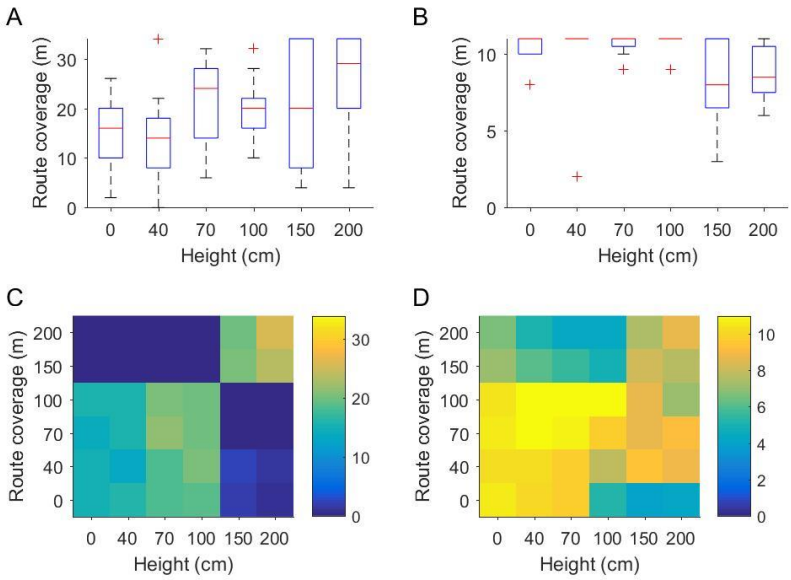

Fig. 7. Rotational catchment areas as a proportion of route length in locations A (left: A and C) and B (right: B and D). X-labels identify the height of the route. A and B show the distribution of rotational catchment areas for routes and goal images at the same height. Red lines indicate the medians, boxes cover the 25th and 75th percentile and whiskers extend to 1.5 times the interquartile range. C and D show rotational catchment areas where the goal (on the x-axis) is at a different height from the route (y-axis). Colours show the mean catchment area across all goal positions in Location A (C) and B (D).

3 Route Navigation

We next test route navigation for images gathered at different heights using the familiarity-based algorithms described in [14]. However, as we need to test the algorithm over a region, we need to use a more controlled robotic platform and move to an indoor gantry robot.

The first version of the algorithm, dubbed 'Perfect Memory' as it stores all training views, proceeds as follows. The agent first traverses a (here pre-defined) route storing greyscale, panoramic images at set distances (represented by red crosses in Fig. 8). Crucially, the stored views are oriented in the direction in which the agent was "facing" at that point. In order to recapitulate the route, the agent then obtains a best-guess of the correct heading by comparing the current and stored views across rotations,

with the rotation yielding the smallest image difference (from among all stored views) taken as the agent's goal bearing. The second method involves the use of an InfoMax familiarity-based neural network [14] and is referred to as Infomax. In this variant, instead of remembering all the views and comparing the current and stored views on an individual basis, a single layer neural network is trained with the stored views to learn the familiarity of the training views using an Infomax training rule [19]. The network is then presented with the current view at a range of rotations and the orientation which yields the smallest activation (and thus the greatest familiarity) is taken as the best-guess heading.

To test these two algorithms, we used an indoor gantry robot, which is comprised of a panoramic camera which outputs a 720x60 pixel image covering 360º of azimuth and 50º of elevation (extending roughly equally above and below the camera's 'horizon'). The camera is mounted onto a robot arm capable of moving along *x*, *y* and *z* dimensions to any arbitrary point within our 2.7 x 1.8 x 1.2 m arena. As this was rather small for our purposes, we treated the arena as though it (and the objects within it) were 10 times as large and scaled the agent's movements appropriately. Henceforth distances in this paper refer to the new artificial scale. For this test we placed a number of cardboard boxes of different heights (min = 0.98m; max = 6.04m) within the arena to provide visual stimuli. Additionally, the walls of the arena were removed so more distal visual cues were available in the form of the visual panorama of the office in which the gantry is housed. First we acquired the stored views for an arbitrary route within the arena at a separation of 25 cm (n = 60; indicated by red crosses in Fig. 8). We then calculated the best-guess headings for the Perfect Memory and InfoMax algorithms (Fig. 8A and B, respectively) for images collected from the gantry at a range of *x*, *y* and *z* coordinates.

Fig. 8 shows that robust performance is given by both the Perfect Memory and InfoMax flavours of the visual homing algorithm. Although performance is understandably poorer for locations where the original training route is obscured by boxes, nonetheless, the algorithms mostly yield headings parallel to the training route, indicating that there is sufficient distal visual information (in the office) to drive reliable homing. Moreover, performance does not decay substantially with changes in height, which suggests that even quite dramatic amounts of visual noise in a real-world environment would be unlikely to lead the agent astray. The quality of performance is particularly remarkable for the InfoMax case, where the only data required in memory is a 644x644 matrix of weights, as opposed to the entire cache of stored views (58 x 720 x 60 pixels in total). Accordingly, the time taken to compute a heading is also considerably faster for InfoMax.

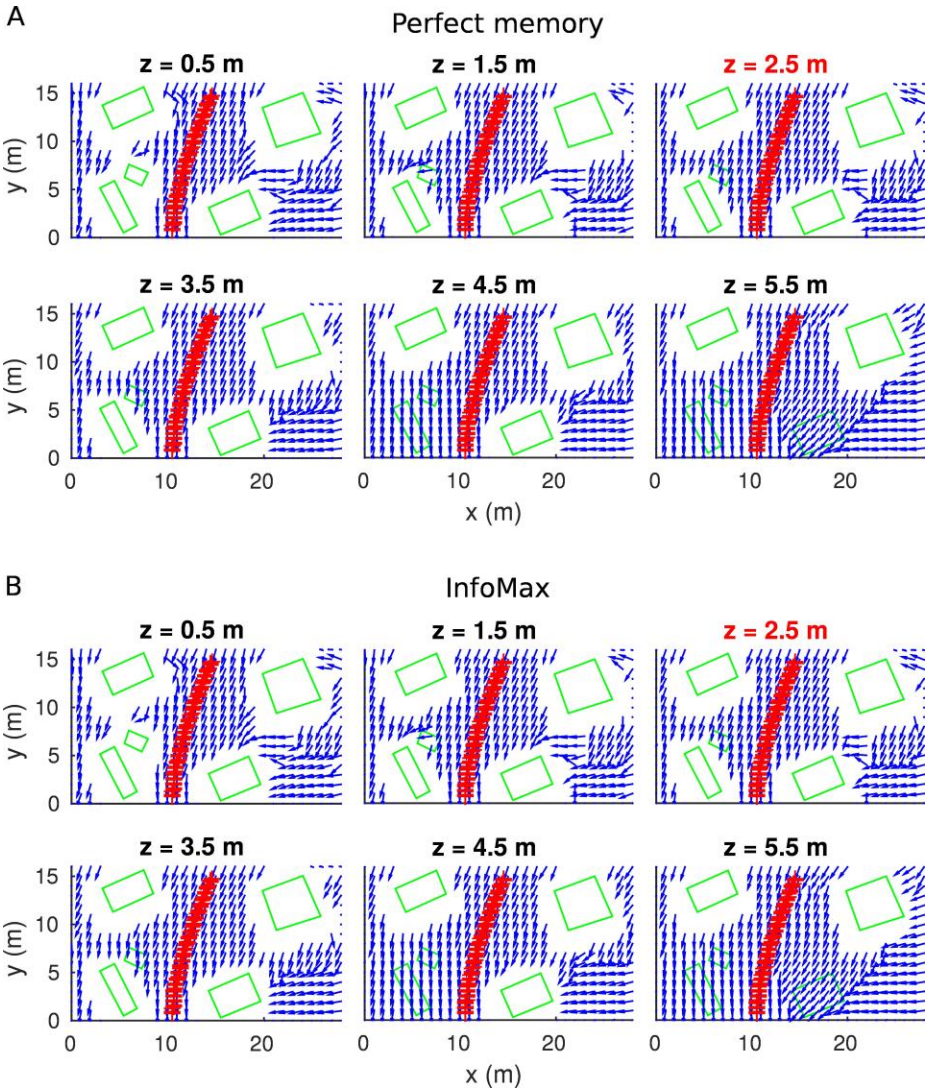

Fig. 8. Estimated headings for different homing algorithms at a range of heights using the robot gantry. **A** – Best-guess headings using the Perfect Memory algorithm. Reliable headings are obtained over a range of locations and the algorithm is robust to changes in height. Blue arrows indicate the headings, red crosses the locations for different stored views and green lines the positions of the cardboard boxes. The scale, as described in the text, is in reality a tenth of that shown in the figure. **B** – Best-guess headings using the InfoMax algorithm. Performance is virtually equivalent to that obtained with the Perfect Memory algorithm, despite requiring far fewer computational resources. Colour scheme is as for A.

4 Conclusions

Here we have shown that parsimonious ant-inspired homing strategies are suitable for aerial robots. Training data collected at one height can be used to robustly recall a heading direction from a range of different heights, indicating that variations in flight height, especially when not too close to the ground, will be tolerated by the algorithm. However, more work needs to be undertaken to prove this performance in closed loop systems, and to test the route algorithm outdoors.

The robustness of performance of visual compass based methods across the range of perspectives also gives hope that route memories could be transferred between robots. The use of a holistic route memory, which is easy to transfer between robots, will aid this endeavor and is the subject of on-going work. In particular, we are interested to see if for instance a UAV could be used to follow a path specified by a ground-based robot, or vice versa. This has implications in, for instance, search and rescue operations where different robots could be used for exploration and retrieval.

Finally, it would be interesting to observe how the flight height of bees vary during foraging. However, currently positional data on bees is generally taken from radar where information on height is unavailable, though efforts are on-going to improve these radar systems which could prove very insightful.

Acknowledgements. Thanks to the anonymous reviewers for helpful suggestions. This work was funded by the Newton Agri-Tech Program RICE PADDY project (no. STDA00732). AP also received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 308943.

References

1. Graham, P., Philippides A: Insect-Inspired Vision and Visually Guided Behavior. In Bhushan B and Winbigler H D (eds.) Encyclopedia of Nanotechnology, Springer (2015)
2. Wehner, R., Räber, F.: Visual Spatial memory in desert ants, Cataglyphis bicolor. Experientia, 35, 1569-1571 (1979)
3. Cartwright, B.A., Collett, T.S.: Landmark Learning in Bees - Experiments and Models. Journal of Comparative Physiology, 151, 521-543 (1979)
4. Wehner, R: Desert ant navigation: How miniature brains solve complex tasks. Karl von Frisch lecture. J Comp Physiol A 189, 579-588 (2003)
5. Wehner, R.: The architecture of the desert ant's navigational toolkit (Hymenoptera: Formicidae). Myrmecol News 12, 85-96 (2009)
6. Collett, T.S., Land, M.F.: Visual spatial memory in a hoverfly. J. Comp. Physiol. A, 100, 59-84 (1975)
7. Junger, W.: Waterstriders (Gerris-Paludum F) Compensate for Drift with a Discontinuously Working Visual Position Servo. J. Comp. Physiol. A. 169, 633-639 (1991).
8. Collett, T.S., Graham, P., Harris, R.A., Hempel-De-Ibarra, N.: Navigational memories in ants and bees: Memory retrieval when selecting and following routes. Advances in the Study of Behavior, 36, 123-172 (2006)

9. Möller, R., Vardy, A.: Local visual homing by matched-filter descent in image distances. Biol Cybern 95: 413-430 (2006)
10. Zeil. J., Hofmann, M., Chahl, J.: Catchment areas of panoramic snapshots in outdoor scenes. J Opt Soc Am A 20, 450-469 (2003)
11. Stürzl, W., Zeil, J.: Depth, contrast and view-based homing in outdoor scenes. Biol Cybern 96, 519-531 (2007)
12. Smith, L., Philippides, A., Graham, P., Baddeley, B., Husbands, P.: Linked local navigation for visual route guidance. Adapt Behav 15, 257–271 (2007)
13. Smith, L., Philippides, A., Graham, P., Husbands, P.: Linked Local Visual Navigation and Robustness to Motor Noise and Route Displacement. In M. Asada et al. (eds.): *SAB 2008, LNCS*, 5040 pp. 179–188. Springer-Verlag, (2008).
14. Baddeley, B., Graham, P., Husbands, P., Philippides, A.: A Model of Ant Route Navigation Driven by Scene Familiarity. PLoS Comput Biol 8(1), e1002336 (2012)
15. Ardin, P., Peng, F., Mangan, M., Lagogiannis, K., Webb, B.: Using an insect mushroom body circuit to encode route memory in complex natural environments. PLoS Comput Biol, 12(2), .e1004683 (2016)
16. Kodzhabashev, A., Mangan, M.: Route Following Without Scanning. In Biomimetic and Biohybrid Systems, pp. 199-210. Springer International Publishing, (2015)
17. Philippides, A., Baddeley, B., Cheng, K., Graham, P.: How might ants use panoramic views for route navigation? J Exp Biol 214, 445-451 (2011)
18. Ardin, P., Mangan, M., Wystrach, A., Webb, B.: How variation in head pitch could affect image matching algorithms for ant navigation. J. Comp. Physiol. A, 201(6), 585-597 (2015)
19. Lulham, A., Bogacz, R., Vogt, S., Brown, M.W.: An infomax algorithm can perform both familiarity discrimination and feature extraction in a single network. Neural Comput 23, 909-926 (2011)